THE ROYAL SOCIETY

Received 26 September 2003
Accepted 21 January 2004
Published online 31 March 2004

Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians

C. Lalueza-Fox[1,2*], M. L. Sampietro[2], M. T. P. Gilbert[3]†, L. Castri[4], F. Facchini[4], D. Pettener[4] and J. Bertranpetit[2]

[1]*Unitat d'Antropologia, Departimenti Biologia Animal, Facultat de Biologia, Universitat de Barcelona, Avinguda Diagonal 645, 08028 Barcelona, Spain*
[2]*Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Dr Aiguader 80, 08003 Barcelona, Spain*
[3]*Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*
[4]*Dipartimento di Biologia e.s., Area di Antropologia, Università di Bologna, Via Selmi 3, 40126 Bologna, Italy*

This study helps to clarify the debate on the Western and Eastern genetic influences in Central Asia. Thirty-six skeletal remains from Kazakhstan (Central Asia), excavated from different sites dating between the fifteenth century BC to the fifth century AD, have been analysed for the hypervariable control region (HVR-I) and haplogroup diagnostic single nucleotide polymorphisms (SNPs) of the mitochondrial DNA genome. Standard authentication criteria for ancient DNA studies, including multiple extractions, cloning of PCR products and independent replication, have been followed. The distribution of east and west Eurasian lineages through time in the region is concordant with the available archaeological information: prior to the thirteenth–seventh century BC, all Kazakh samples belong to European lineages; while later an arrival of east Eurasian sequences that coexisted with the previous west Eurasian genetic substratum can be detected. The presence of an ancient genetic substratum of European origin in West Asia may be related to the discovery of ancient mummies with European features in Xinjiang and to the existence of an extinct Indo-European language, Tocharian. This study demonstrates the usefulness of the ancient DNA in unravelling complex patterns of past human migrations so as to help decipher the origin of present-day admixed populations.

Keywords: Central Asia; past migrations; mitochondrial DNA; ancient DNA

1. INTRODUCTION

As an area of contact between Asia and Europe, Central Asia has experienced a scenario of complex cultural developments, large migratory movements and biological admixture between populations from west Eurasia and East Asia (Bowles 1977). The role of the region in the spread of modern humans is not clear, however. One hypothesis envisages the region's role as a corridor towards the East, as depicted in many representations of the Out of Africa hypothesis. By contrast, genetic evidence points to a late settlement of already differentiated East and West modern human populations (Comas *et al.* 1998; Calafell *et al.* 2000).

The vast steppes of Central Asia are particularly suitable for pastoral nomadism. Pastoralist groups are extremely mobile and their way of life includes periods of rapid population growth associated with periodic increments in the size of the herds. In the second century BC, one such demographic explosion occurred in the steppes that provoked the first of a series of migrations into Central Asia. The first migration consisted of a group of pastoral nomads named the Sakas (or Scythians), who departed from the European area of the steppes and spread east into Asia (Ismagulov 1970; Ismagulov *et al.* 1998). Later migrations, however, spread from the east toward the west. These included the Hsiung-nu, referred to in the Chinese annals and generally considered as the ancestors of the Huns, around the seventh century BC (Ismagulov 1982; Ismagulov *et al.* 1998).

Recently, hundreds of mummified humans have been discovered in the Tarin Bassin area of Xinjiang, west China, which date between 2000 BP and AD 200–300. These mummies are notable for their European features, (e.g. blond or reddish hair, and straight noses), and reveal an ancient Caucasoid substratum that was present far into the east of Asia (Mair 1995). However, Hemphill (1999) and Hemphill & Mallory (2004) have also noted cranial similarities of some Tarim mummies to populations of South Asia. Ancient mtDNA analysis of some of these remains (Francalacci 1995, 1998) has also demonstrated that they contained a European lineage. It has been suggested that the presence of these individuals in this area is related to the existence of an isolated Indo-European language, Tocharian (Hamp 1998), that seems to have affinities with Armenian, Thracian, Phrygian, Hittite and Celtic (Ruhlen 1996) and that survived in this area until the beginning of the second century AD (Ruhlen 1996).

Several recent genetic studies have analysed classical markers of populations of Central Asia, such as blood groups (Fiori *et al.* 2000), mtDNA (Comas *et al.* 1998)

* Author for correspondence (carles.lalueza@upf.edu).
† Present address: Department of Ecology and Evolutionary Biology, BSW 222, University of Arizona, 1041 East Lowell Street, Tucson, AZ 85721, USA.

Table 1. Sample distribution by areas, site, age and period.

area	sites	period	age (centuries)	samples (code)
central	Ak-Mustafa (Kourgan 2)	Bronze	thirteenth to eleventh BC	1, 4, 5
	Birlik (Kourgan 1,8,9,11,12,23,29)	Early Iron	eighth to fifth BC	9, 10,
	Birlik (Kourgan 15)	Iron	seventh to third BC	7, 8, 13, 14, 17, 18, 19
south	Oi-Zhaylau-III, Talapty-II	Bronze	fourteenth to eleventh BC	24, 25
	Molaly (Kourgan 14,18,20,25)	Iron	seventh to fifth BC	89, 95
	Shubarat (Kourgan 9,32,33,81)			
	Kok-Mardan (Kourgan 2,3,6,16)	Antique	third to fifth AD	28
east	Izmaylovka (Kourgan 7,10)	Bronze	thirteenth to eleventh BC	41, 42, 85, 88
	Izmaylovka (stone fencing)			
	Zevakinskiy (stone fencing)	Bronze/Iron	eighth to seventh BC	47, 48, 52, 55, 56
west	Vodokhranilische (Kourgan 4)	Bronze	fourteenth to tenth BC	57, 60, 61, 62, 84
	Rybniy Sakryl-III (Kourgan 3,5,6,7,8,10)			
	Solyanka II (Kourgan 2,3)	Late Iron	fifth to first BC	66, 68, 74, 78, 79
	Solyanka I (Kourgan 1)			
	Solyanka-Limany (Kourgan 8)			
	Kos-Oba (Kourgan 8,9,16)			
	Rybniy-Sakryl II (Kourgan 2)			
	Shaytan-Oba (Kourgan 2)			

and Y-chromosome variability (Perez-Lezaun *et al.* 1999; Wells *et al.* 2001; Zerjal *et al.* 2002). These studies have all demonstrated that the present-day populations of Central Asia are genetically extremely heterogeneous, with the presence of well-differentiated east Eurasian and west Eurasian lineages in all populations studied.

Various ancient DNA studies have also attempted to reconstruct population movements in Asia, and in particular in China and Mongolia (Oota *et al.* 1999; Wang *et al.* 2000; Keyser-Tracqui *et al.* 2003). The results of two of these studies initially suggested that 2500 years ago the human population from Linzi, Shandong province, had genetic affinities with present Europeans (Wang *et al.* 2000). However, a reanalysis of the Linzi sequences in the light of the HVS-I (hypervariable control region 1 of the mtDNA) substitutions associated to specific mtDNA haplogroups (Yao *et al.* 2003), demonstrates the opposite—that the ancient population did not have any specific affinity to the European gene pool, and instead is related to modern southern Chinese populations.

In this study, we have determined the mtDNA sequences and haplogroups from several prehistoric samples from Kazakhstan, a core place within the landscape of Central Asia, to help unravel some of the early steppe migrations. The samples span different time periods and cover both the eastern and western regions. To avoid ambiguous conclusions (such as those seen in the previous Asian ancient DNA studies) we have followed the suggestions of Yao *et al.* (2003) and use several different mtDNA coding-region single nucleotide polymorphism (SNP) markers (Richards *et al.* 1996, 2000; Macaulay *et al.* 1999; Ingman *et al.* 2000; Kivisild *et al.* 2002; Herrnstadt *et al.* 2002) to corroborate the haplogroups suggested by the HVS-I sequences.

2. MATERIAL AND METHODS

Thirty-six teeth retrieved from Kazakh archaeological sites (table 1) and belonging to different individuals were chosen by D.P. on the basis of their macroscopic preservation and subjected to a DNA analysis at the ancient-DNA laboratory in Barcelona. A subsample of six teeth was also sent directly from Bologna to the Ancient Biomolecules Centre in Oxford to enable independent replication of the DNA results. DNA was extracted from the samples as published in detail elsewhere (Lalueza-Fox *et al.* 2001). In brief, teeth were powdered then decalcified overnight in 10 ml of 0.5 M EDTA. Subsequently, samples were incubated overnight in a digestion mix consisting of 1 ml of 10% SDS, 0.5 ml of 1 M Tris–HCl and 100 µl of 1 mg ml^{-1} proteinase K. Post-digestion, DNA was extracted using phenol/chloroform and desalted using Centricon 30 micro-concentrators (Amicon). In accordance with suggested ancient DNA procedures (Cooper & Poinar 2000) the DNA extractions were performed in an isolated pre-PCR area that is dedicated to ancient DNA, and contains both positive air pressure and overnight UV-light exposure. Blank extraction and amplification controls were incorporated throughout the analysis, to monitor for the presence of possible contamination. Sterile aliquoted reagents, sterile gloves, face masks, filter pipette tips, frequent surface bleaching and other standard precautions of ancient DNA studies were adopted.

PCR amplifications were performed in 25 µl reactions with 1 µl of extract, 1 U EcoTaq and 1× buffer (EcoGen), 2 mg ml^{-1} BSA, 2.5 mM $MgCl_2$, 0.25 mM dNTPs and 1 µM each primer; the PCR profile was 40 cycles at 94 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min, with a first denaturation step of 94 °C for 5 min. PCR products were visualized in a 0.8% low-melting-point agarose gel, and the appropriate bands were excised from the gel, melted in sterile water and subjected to a second PCR. PCR products were purified with columns (Qiagen) and sequenced with BigDye 3.1 in an ABI 3100 DNA sequencer (Applied Biosystems). Cloning of the PCR products was performed with SureClone Ligation Kit (Pharmacia, Uppsala, Sweden) following the supplier's instructions. Extraction, amplification, cloning and sequencing in the Oxford laboratory followed slightly different conditions (Gilbert *et al.* 2003*a*). Various combinations of previously published primers (Handt *et al.* 1996) were used to amplify the mitochondrial

Table 2. Primers and markers used for haplogroup attribution (L16517 and H160 were designed by Agnar Helgason, *deCODE* Genetics).

	haplogroup	marker position
L635 (5′-TGAAAATGTTTAGACGGGCTCACATC-3′) H677 (5′-GCATGTGTAATCTTACTAAGAG-3′)	A haplogroup	*Hae*III 663 (+)
L4818 (5′-CTTTCACTTCTGAGTCCCAG-3′) H4851 (5′-TTGTCATGTGAGAAGAAGCAG-3′)	G haplogroup	*Hha*I 4 831 (+)
L6999 (5′-CAAACTCATCACTAGACATCG-3′) H7066 (5′-GAATGAAGCCTCCTATGATGG-3′)	H haplogroup	*Alu*I 7 025 (−)
L16517 (5′-CATCTGGTTCCTACTTCAGG-3′) H160 (5′-ATCGCACCTACGTTCAATATTACA-3′)	HV haplogroup	np 73
L10014 (5′-TTTTAGTATAAATAGTACCG-3′) H10088 (5′-GTAGTAAGGCTAGGAGGGTG-3′)	I haplogroup	*Alu*I 10 028 (+)
L10355 (5′-TATTAATCATCATCCTAGCC-3′) H10426 (5′-TTAATGAGTCGAAATCATTC-3′)	M haplogroup	np 10 398–10 400
L13257 (5′-AATCGTAGCCTTCTCCACTTCA-3′) H13393 (5′-TCCTATTTTTCGAATATCTTGTTC-3′)	T haplogroup	*Bam*HI 13 366 (+)
L12257 (5′-CCCATGTCTAACAACATGG-3′) H12341 (5′-GGTTATAGTAGTGTGCATGG-3′)	U haplogroup	*Hinf*I 12 308 (+)
L8215 (5′-ACAGTTTCATGCCCATCGTC-3′) H8297 (5′-ATGCTAAGTTAGCTTTACAG-3′)	W haplogroup	*Ava*II 8 249 (+)

HVS-I (as required by the individual samples). The primers used were L16 022, L16 055, L16 122, L16 185, L16 209, H16 131, H16 142, H16 218, H16 356 and H16 401.

As a result of its relatively high evolutionary rate, it is both difficult and inadvisable to assign unique mtDNA haplogroups to specimens using control-region DNA alone (Richards *et al.* 1996; Macaulay *et al.* 1999, among others). Thus, various diagnostic SNP markers were chosen to aid haplogroup determination as displayed in table 2.

Sequence electropherograms that indicated heteroplasmic positions, attributable to damage or contamination *post mortem* (Gilbert *et al.* 2003*b*), were either repeated and/or cloned. If an observed substitution is due to a random DNA damage and Taq misincorporation, it is expected that this will not appear in the same position in another amplification from the same extract. Therefore, most of the samples were amplified and sequenced several times, using different sets of primers. In Barcelona, samples with discrepant results were cloned; in Oxford, all samples were cloned.

3. RESULTS

Twenty-nine out of the 36 samples yielded amplifiable DNA. This high rate of success suggests that the DNA in the samples is well preserved, as can be expected in samples that originate from cold and arid environmental conditions such as those found at most of the Kazakhstan sites. Further evidence of the well-preserved state of the DNA is suggested from preliminary assays that indicate that single-copy nuclear genes can be retrieved from *ca.* 10% of the samples.

It was possible to estimate a damage rate (number of observed substitutions per 1000 bp) from the results of 106 different clones (Cooper *et al.* 2001). The values ranged from 0 to 2.68%, rather a low figure, similar to those observed in other DNA studies and close to modern DNA estimates (Cooper *et al.* 2001; Hofreiter *et al.* 2001; Gilbert *et al.* 2003*b*).

In addition to the well-preserved nature of the DNA, evidence for the authenticity and accuracy of the retrieved sequences is offered through the consistent nature of the sequence retrieval throughout both independent DNA extractions and amplification of the DNA using different

Table 3. Amplification strategy of the analysed samples. (*Sample independently replicated; C, cloned; D, samples discarded owing to irreproducible results or mixed contamination; E, multiple extractions from the same individual.)

sample	additional typing	HVS-I fragments sequenced
14	C	055–218/209–401
17	E	055–131/055–218/209–401/185–356
8		055–218/185–401
62	E	055–142/055–218/209–401
74		055–218/209–401
85	*/C	055–142/209–401/185–356
47		055–401/209–401
56		022–218/185–401
48	E	055–142/055–401/131–218/209–401
9	*/C	055–218/209–401/185–356
18	E	022–218/055–142/055–401/185–401/185–356
66		022–218/185–401/185–356
78		022–218/185–401/185–356
60		055–218/185–401
89	E	055–142/055–218/185–401/209–401
19	C	022–218/185–401
95		022–218/055–401/185–401
84		055–218/185–356
25		055–218/209–401
24		055–218/185–401
13		022–218/185–401
28		055–218/209–401
52		022–218/185–401
1	*/C	055–218/055–401/209–401
41		055–218/185–401
55	*/C	022–218/185–401
10		055–218/209–401/185–356
68	*/C/D	055–218/209–401
79	*/C/D	022–218/185–401/185–356

Table 4. Haplogroup attribution and HVR1 haplotype for each sample successfully analysed.

haplogroup	sample	HVS-I haplotype	SNPs
east Eurasian ($n = 6$)			
A*	8, 17	223 319 362	+663*Hae*III
A*	14	223 290 319 362	+663*Hae*III
M*	19	223 286	10398G/10400T
M4	89	129 174 223	10398G/10400T
G2	10	189 223 278	+4831 *Hha*I
west Eurasian ($n = 21$)			
H	9, 47, 48, 56, 85	CRS	−7025 *Alu*I
H	18	266 287	−7025 *Alu*I
H	66	150 189	−7025 *Alu*I
HV	62, 74	129	+7025 *Alu*I, 73A
I	60	129 223 294	+10032 *Alu*I
T*	95	126 294 296	+13366 *Bam*HI
T*	84	126 288 294 296	+13366 *Bam*HI
T1	24	126 163 186 189 294	+13366 *Bam*HI
T1	25	093 126 163 186 189 294	+13366 *Bam*HI
U*	13	189	+12308 *Hinf*I
U*	78	311	+7025 *Alu*I, 73G
U1	28	183 189 249	+12308 *Hinf*I
U1	52	189 249	+12308 *Hinf*I
U5	1	187 192 260 270	+12308 *Hinf*I
U5a1	41	192 256 270 311	+12308 *Hinf*I
W	55	223 292	+8249 *Ava*II

primer sets (table 4). The concordance of key SNPs typed along the coding region with the HVR-I sequences (table 4) lends further support to the mtDNA haplotypes found. Also, in five specimens multiple extractions from different teeth (table 3) always yielded consistent sequences.

Four samples replicated in Oxford yielded the same sequence as those found in Barcelona and proved the independent reproducibility of the results generated. However, two samples were discarded: one because of mixing of sequences in both laboratories, and the other because of inconsistency (the haplotype retrieved in Barcelona by direct sequencing was T16 223-C16 298; but when the PCR products from this sample were cloned in Oxford, the generated sequences showed that in fact it was the result of an unusual mixture of three sequences (one with T16 223, one with T16 278 and another with C16 298) that produced a false, compound sequence of a putative C haplogroup with no possibility to recognize which was the endogenous sequence).

Most of the retrieved sequences ($n = 21$, 78%) belong to European or west Eurasian mtDNA haplogroups (HV, H, T, I, U and W haplogroups). HV lineages are represented by two sequences with a 16,129 substitution, commonly found in the central Mediterranean regions (Richards *et al.* 2000). The most frequent H haplotype observed in our sample is CRS; the other two H sequences, characterized by 16 266/16 287 and 16 150/16 189 motifs, are also present in the central Mediterranean and in the Caucasus, respectively (Richards *et al.* 2000). Haplogroup I is present in one individual from Central Asia with a motif reported also in some individuals from the Caucasus (Richards *et al.* 2000), while the haplogroup W sequence is widespread in west Eurasia (Richards *et al.* 1996, 2000). Ancient members of T haplogroups of Central Asia can be further subdivided into subgroups T* and T1. Sequences T* with the same motifs as those reported for ancient populations from Central Asia can be found widespread in Europe and the Middle East (16 126/16 294/16 296) and in the central Mediterranean area (16 126/16 288/16 294/16 296). T1 is present, along with the root sequence (16 126/16 163/16 186/16 189/16 294), is widespread in western Eurasia but with occasional occurrences in eastern Eurasia, and one derivative with an additional mutation in 16 093, typical of populations in Central Asia and Turks (Comas *et al.* 1998; Richards *et al.* 2000; L. Castri, P. Francalacci, L. Simoni, D. Luiselli, F. Facchini and D. Pettener, unpublished data). Haplogroup U is represented by six sequences belonging to subgroups U*, U1 and U5. U1a sequences with a 16 189/16 249 motif can be found also in Turks, Armenians and Caucasians. Sequence U5a has been described in only one Egyptian (Richards *et al.* 2000), whereas sequence U5a1 is frequent in the Caucasus, but present also in Europe; a different U5a1 haplotype has also been found in ancient samples from Mongolia (Keyser-Tracqui *et al.* 2003).

The eastern Eurasian lineages (22% of the sequences) are represented by sequences belonging to haplogroups M, G and A. The haplogroup M sequences can be further assigned to subgroups M*, M4 (Kivisild *et al.* 1999) and G2. The M* sequence has an additional mutation at 16,286, which has previously been observed in only one Indian sample (Kivisild *et al.* 1999), whereas the M4 sequence contains a T at 16 174 that has never previously been reported in other populations. G2 (HVRI motif 16 223/16 278) is present in modern populations in Central Asia and China, but the ancient central Asian sequence has an additional mutation at 16 189. Haplogroup A is represented by three sequences: one has a HVRI motif commonly found in Central Asia (Tuvinians,

Table 5. West Eurasian and east Eurasian samples typed for each time period in different geographical areas from Kazakhstan. (Numbers refer to western Eurasian sequences and to total sequences found. An asterisk denotes an Indian sequence included with eastern Eurasian.)

	fourteenth to tenth century BC	eighth to seventh century BC	seventh to third century BC	third to fifth century AD
east	2/2	5/5	—	—
south	2/2	—	1*/2	1/1
central	1/1	—	3/8	—
west	3/3	—	3/3	—
total	8/8	5/5	7/13	1/1

Kazaks, Kirghizs and Mongolians) and India, while the remaining two have a retromutation at nucleotide position 16 290; the same motif has been found in one Siberian Chukchi (Starikovskaya *et al.* 1998).

The sequences from populations from Central Asia from this study display a marked temporal pattern that is concordant with the available archaeological information (table 5). Before the seventh century BC there is a conspicuous absence of haplogroups in East Asia (although it can be estimated that haplogroups from East Asia could be present up to a frequency of 20.6% and not been detected with the analysis of only 13 individuals, $p < 0.05$); after that period, samples consist of a mixture of east and west Eurasian sequences, with almost half of them belonging to East Asian haplogroups (6 out of 14 individuals; 43%). The first detection of east Eurasian haplogroups varies between geographical areas. No east Eurasian lineages appear in east Kazakhstan samples that date to the eighth to seventh century BC or earlier. In south Kazakhstan, the first Asian lineages appear in the seventh to fifth century BC. In central Kazakhstan, east Eurasian lineages are present in samples from the eighth to third century BC, and in the seventh century BC. In west Kazakhstan, even samples dating to the most recent period examined do not contain any east Eurasian sequences.

4. DISCUSSION

mtDNA is particularly suited to the study of admixed populations such as those of Central Asia: the genealogy is now well established and there is a clear geographical structuring of the east and west Eurasian lineages. Moreover, the past peopling of Central Asia can only be really understood with the analysis of ancient material from different periods and the unambiguous haplogroup attribution, owing to the complexity and superimposition of the migration patterns involved.

The ancient Kazakh samples cluster in nodes already detected in the general mtDNA network of Central Asia (figure 1). Mitochondrial haplotypes among modern Kazakhs have been reported in two separate studies as follows: 45–31.7% containing haplotypes commonly found in east Eurasian; 0–5% containing haplotypes commonly found in India; and 50–63.4% containing haplotypes commonly found in west Eurasia (Comas *et al.* 2004; L. Castri, P. Francalacci, L. Simoni, D. Luiselli, F. Facchini and D. Pettener, unpublished data, respectively). A yet smaller fraction of sequences (0–4.9%) corresponds to haplogroups such as N and R that are widely distributed along Eurasia. The general west–east Eurasian composition of the prehistoric samples in the period after the arrival of east Eurasian sequences (after the seventh century BC) is, despite the small sample size ($n = 14$), quite similar to the values found in the modern Kazakh population: east Eurasian (42.9%), west Eurasian (50%) and Indian (7.1%). Interestingly, the only sequence of Indian origin that was observed, belonging to the M4 haplogroup (Bamshad *et al.* 2001), originates from a site in the south of Kazakhstan. This fact could correspond to an independent, Indo-Iranian genetic infusion into the steppes. Another sequence belonging to the M4 sub-haplogroup has been found in present-day Kazakhs, although other populations of Central Asia, such as the Uzbeks, Kyrgyz, Tajiks, Turkmen and Uighurs seem to lack this sub-haplogroup (Comas *et al.* 2004).

The haplogroup composition differs slightly between modern Kazakhs and prehistoric samples that originate from prior to the seventh century BC, with variations attributable to time structure and differences in sample sizes (table 5). Haplogroups present in modern Kazakhs, such as B, F, C, Z, D, R, J and Y, were not observed in the prehistoric Kazakhs. By contrast, two haplogroups observed among the ancient samples, W and I, have not yet been found among modern Kazakhs. The results also indicate that there is an excess of west Eurasian haplogroups in comparison with those currently found (notably haplogroups H and U). However, this may be attributed to the overrepresentation of the earlier temporal period with only west Eurasian haplogroups. The observed absence of east Eurasian sequences prior to the eighth to seventh century BC suggests an earlier prehistoric expansion of peoples containing west Eurasian sequences into Asia, that probably went further east, into present-day China. This expansion may be related to the discovery of mummies that contain European features and west Eurasian mtDNA sequences in the Tarin basin, China, as well as the relict Indo-European Tocharian. An intriguing finding is the occurrence in the ancient sample of sequences from Central Asia that are mainly distributed far into the West, such as the Caucasus and central Mediterranean areas.

The latter westward movement of Asian nomads, such as that of the Hsiung-nu, estimated to have taken place around the sixth century BC (Ismagulov 1982), corresponds to the period in which the earliest east Eurasian sequences are found among our ancient samples. In addition, the presence of a haplogroup A sequence found in Siberia (Starikovskaya *et al.* 1998) and a G2 sequence found in Chinese Han (Yao *et al.* 2002) points to Siberia and Mongolia as a possible source of such migrations.

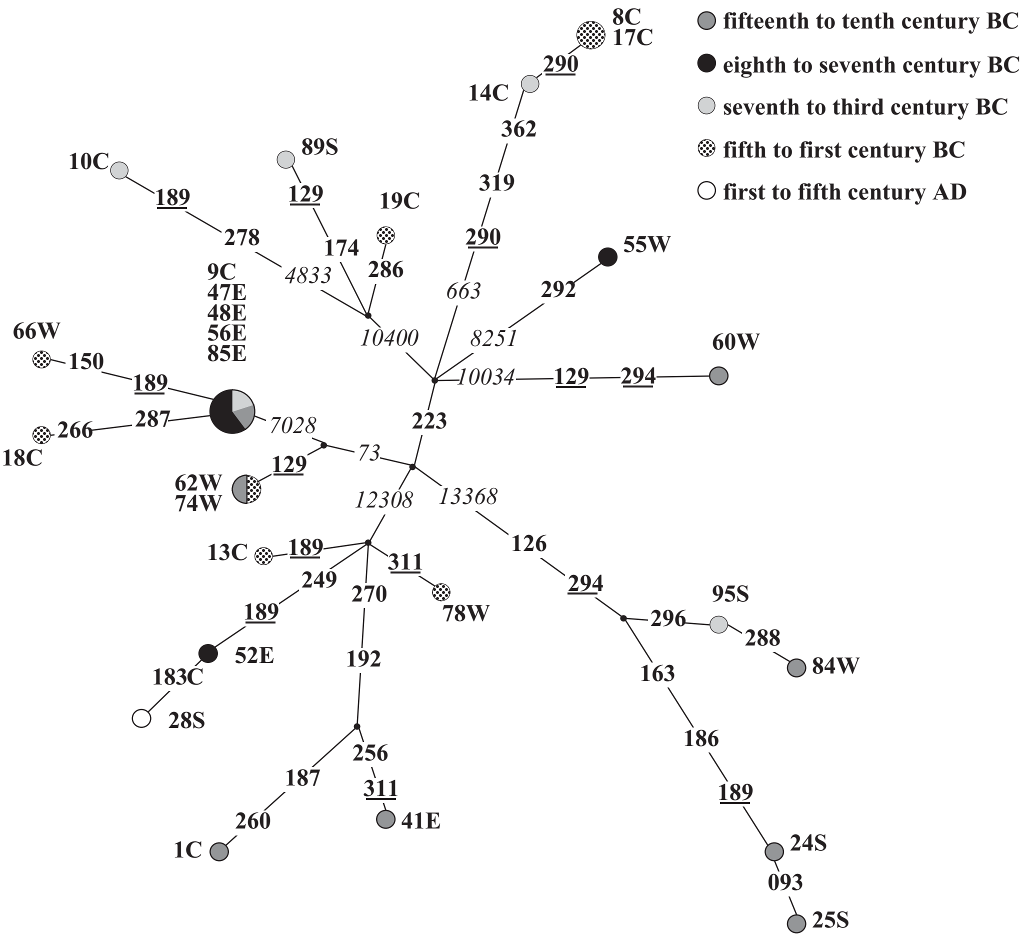

Figure 1. Reduced median network of Central Asia sequences, both modern (data from L. Castri, P. Francalacci, L. Simoni, D. Luiselli, F. Facchini and D. Pettener, unpublished data) and ancient. Node size is proportional to frequencies. HVR1 positions are numbered relative to CRS (minus 16 000); RFLPs diagnostic positions are in italics. Recurrent mutations are underlined. Ancient samples are numbered according to table 4; the suffix refers to the geographical locations of ancient Kazakh samples (central, east, west, south), and the circle tones refer to the different archaeological periods (see key).

If there was a human substratum of west Eurasian origin far into East Asia in prehistoric times, what happened to it? Analyses of present-day Han mtDNA sequences from different regions in China detect a very residual presence (less than 5%) of European haplotypes in a few regions. These include Qinghai (east to Xinjiang and Tibet) and Yunnan (north of India and Burma), as well as some coastal regions (Yao *et al.* 2002). Moreover, in a sample from the region adjacent to Kazakhstan, Xinjiang, no European lineages have been found among the 47 individuals sampled so far (Yao *et al.* 2002).

In contrast to the methodological pitfalls of several previous ancient DNA studies in Asia, we have followed the same haplogroup designation guidelines proposed by Yao *et al.* (2003). Therefore, the results are fully comparable, and it is interesting that the only two west Eurasian lineages found in modern Chinese correspond to T1 and HV haplogroups (Yao *et al.* 2003), are also found among the prehistoric Kazakhs. This suggests that the genetic legacy of the prehistoric European eastward movement was erased by later Asian expansions, and thus had almost no genetic contribution to the present-day Asians. Such extinction may be related to the demographic processes that also lead to the disappearance of the Tocharian language.

Many more samples, sites and populations will need to be studied to clarify the prehistory of Central Asia. Nevertheless, this study demonstrates that ancient DNA can be a useful tool to help interpret contemporary genetic data in admixtured populations, through the reconstruction of past migrations.

This research was supported by the Dirección General de Investigación, Ministerio de Ciencia y Tecnología of Spain (grants BOS2001-0794 and BMC2001-0772), by Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (grant 2001SGR00285) and a fellowship AP2002-1065 to M.L.S. M.T.P.G. is grateful to the Wellcome Trust for funding his research. The authors are grateful to Orazak Ismagulov and Ainagul Ismagulova (Department of Anthropology, Central State Museum of Republic of Kazakhstan) for providing them with the ancient samples, and to David Comas for his comments on the manuscript.

REFERENCES

Bamshad, M. (and 17 others) 2001 Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11**, 994–1004.

Bowles, G. T. 1977 *The people of Asia*. Birkenhead, UK: Willmer Brothers Limited.

Calafell, F., Comas, D., Pérez-Lezaun, A. & Bertranpetit, J. 2000 Genetics and population history of Central Asia. In *Archaeogenetics: DNA and the population prehistory of Europe* (ed. C. Renfrew & K. Boyle), pp. 259–266. Cambridge: McDonald Institute for Archaeological Research.

Comas, D. (and 11 others) 1998 Trading genes along the Silk Road: mitochondrial DNA sequences and the origin of Central Asian populations. *Am. J. Hum. Genet.* **63**, 1824–1838.

Comas, D., Plaza, S., Wells, R. S., Yuldusheva, N., Lao, O., Calafell, F. & Bertranpetit, J. 2004 Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* (In the press.)

Cooper, A. & Poinar, H. N. 2000 Ancient DNA: do it right or not at all. *Science* **289**, 1139.

Cooper, A., Lalueza-Fox, C., Anderson, S., Austin, J., Rambaut, A. & Ward, R. 2001 Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704–707.

Fiori, G., Facchini, F., Ismagulov, O., Ismagulova, A., Tarazona-Santos, E. & Pettener, D. 2000 Lung volume, chest size, and hematological variation in low-, medium-, and high-altitude central Asian populations. *Am. J. Phys. Anthropol.* **113**, 47–59.

Francalacci, P. 1995 DNA analysis of ancient desiccated corpses from Xinjiang. *J. Indoeur. Stud.* **23**, 385–389.

Francalacci, P. 1998 DNA analysis on ancient desiccated corpses from Xinjiang (China): further results. In *The Bronze Age and early Iron Age peoples of eastern Central Asia* (ed. V. H. Mair), pp. 537–547. Washington, DC: Institute for the Study of Man.

Gilbert, M. T., Hansen, A. J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N. & Cooper, A. 2003*a* Characterization of genetic miscoding lesions caused by postmortem damage. *Am. J. Hum. Genet.* **72**, 48–61.

Gilbert, M. T., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N. & Cooper, A. 2003*b* Distribution patterns of post mortem damage in human mitochondrial DNA. *Am. J. Hum. Genet.* **72**, 32–47.

Hamp, E. P. 1998 Whose were the Tocharians?—Linguistic subgrouping and diagnostic idiosyncrasy. In *The Bronze Age and early Iron Age peoples of eastern Central Asia* (ed. V. H. Mair), pp. 307–346. Washington, DC: Institute for the Study of Man.

Handt, O., Krings, M., Ward, R. H. & Pääbo, S. 1996 The retrieval of ancient human DNA sequences. *Am. J. Hum. Genet.* **59**, 368–376.

Hemphill, B. E. 1999 Biological affinities and adaptations of Bronze Age Bactrians: IV. A craniometric investigation of Bactrian origins. *Am. J. Physiol. Anthropol.* **108**, 173–192.

Hemphill, B. E. & Mallory, J. P. 2004 Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from western Central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *Am. J. Physiol. Anthropol.* (In the press.)

Herrnstadt, C. (and 10 others) 2002 Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian and European haplogroups. *Am. J. Hum. Genet.* **70**, 1152–1171.

Hofreiter, M., Jaenicke, V., Serre, D., Von Haeseler, A. & Pääbo, S. 2001 DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acid Res.* **29**, 4793–4799.

Ingman, M., Kaessmann, H., Pääbo, S. & Gyllenstein, U. 2000 Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713.

Ismagulov, O. 1970 *The population of Kazakhstan from the epoch of Bronze up to modern time*. Alma Ata: Institute of History, Archaeology and Ethnography of the Academy of Science of Kazakhstan.

Ismagulov, O. 1982 *Ethnic anthropology of Kazakhstan*. Alma Ata: Institute of History, Archeaology and Ethnography of the Academy of Science of Kazakhstan.

Ismagulov, O., Sikhymbaeva, K. & Ismagulova, A. 1998 *History of Kazakhstan*. Alma Ata: Gylym.

Keyser-Tracqui, C., Crubezy, E. & Ludes, B. 2003 Nuclear and mitochondrial DNA analysis of a 2000-year-old necropolis in the Egyin gol valley of Mongolia. *Am. J. Hum. Genet.* **73**, 247–260.

Kivisild, T. (and 14 others) 1999 Deep common ancestry of Indian and Western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **9**, 1331–1334.

Kivisild, T., Tolk, H. V., Parik, J., Wang, Y., Papiha, S. S., Bandelt, H. J. & Villems, R. 2002 The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* **19**, 1737–1751.

Lalueza-Fox, C., Luna-Calderón, F., Calafell, F., Morera, B. & Bertranpetit, J. 2001 MtDNA from extinct Tainos and the peopling of the Caribbean. *Ann. Hum. Genet.* **65**, 137–151.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. & Torroni, A. 1999 The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**, 232–249.

Mair, V. H. 1995 Prehistoric Caucasoid corpses of the Tarim Bassin. *J Indoeur. Stud.* **23**, 281–370.

Oota, H., Saitou, N., Matsushita, T. & Ueda, S. 1999 Molecular genetic analysis of remains of a 2000-year-old human population in China—and its relevance for the origin of the modern Japanese population. *Am. J. Hum. Genet.* **64**, 250–258.

Perez-Lezaun, A. (and 11 others) 1999 Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* **65**, 208–219.

Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J. & Sykes, B. 1996 Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**, 185–203.

Richards, M. (and 36 others) 2000 Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276.

Ruhlen, M. 1996 *On the origin of languages: studies on linguistic taxonomy*. Stanford, CA: Stanford University Press.

Starikovskaya, Y. B., Sukernik, R. I., Schurr, T. G., Kogelnik, A. M. & Wallace, D. C. 1998 mtDNA diversity in Chukchi and Siberian eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am. J. Hum. Genet.* **63**, 1473–1491.

Wang, L., Oota, H., Saitou, N., Jin, F., Matsushita, T. & Ueda, S. 2000 Genetic structure of a 2500-year-old human population in China and its spatiotemporal changes. *Mol. Biol. Evol.* **17**, 1396–1400.

Wells, R. S. (and 26 others) 2001 The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl Acad. Sci. USA* **98**, 10 244–10 249.

Yao, Y. G., Kong, Q. P., Bandelt, H. J., Kivisild, T. & Zhang, Y. P. 2002 Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* **70**, 635–651.

Yao, Y. G., Kong, Q. P., Man, X. Y., Bandelt, H. J. & Zhang, Y. P. 2003 Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol. Biol. Evol.* **20**, 214–219.

Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R. & Tyler-Smith, C. 2002 A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am. J. Hum. Genet.* **71**, 466–482.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.